This document was produced by SDC in performance of DSA-600-12301

TM-WD-268/004/00
AUTHOR R. Watson
TECHNICAL K. Yung
RELEASE K. Yung
for
DATE 19 Sep 60 PAGE 1 OF 3+ PAGE

TECH MEMO

a working paper

System Development Corporation / 5720 Columbia Pike / Falls Church, Virginia 22041

DEFENSE DOCUMENTATION CENTER FIVE YEAR PLAN STUDY

APPENDIX

This Appendix contains recommendations for ADP-related developments in the next five years. Many of the developments are listed or implied under functional headings in other sections of the report. For the sake of clarity they are listed here in a single section. Some of these developments are expected to extend beyond the five year period.

1. ENVIRONMENTAL CONSIDERATIONS FOR TOTAL DDC COMPUTER DEVELOPMENT AND RESEARCH

The computer-oriented research and development to be detailed in this Appendix falls within the environment of an integrated internal systems design outlined in the main body of the report. The report discusses this system effort as encompassing the following five year period; this appendix will consider the integrated system effort as extending beyond that period and as the contextual vehicle for extensive and advanced computer-oriented research and development. The integrated system design, as outlined in the main report body, essentially is concerned with internal DDC operations and with the current and foreseeably expanded group of corporate and individual users of DDC products and services. It is also concerned, external to DDC per se, with inter-Federal

WASHINGTON OPERATIONS CENTER
CORPORATE OFFICES: Santa Monica, California

agency agreements on the servicing of user requests; with the provision (on the part of several agencies) of materials going into one given mission-oriented area class of product or service, and with standards and formats for documentary data used in interagency transmission. This Appendix to the SDC report will consider this interagency environment as well as the immediate environment of DDC and its users.

Let us proceed to a discussion of various facets of equipment development which may have long range applicability for DDC. We will consider both software (programs and groups of programs) and equipment. Within this context we will consider DDC Computer Center research and development activities as being confined to the generic application areas found in performing the tasks, now and in the foreseeable future, assigned to the Defense Documentation Center.

2. COMPUTER SOFTWARE AND PROGRAM DEVELOPMENT ACTIVITIES FOR THE DEFENSE DOCUMENTATION CENTER

The overriding need here is selection of one type of computer within the confines of the integrated system development effort. This type of machine should be compatible across the industry insofar as is possible, to enable most efficient and rapid transfer of data, processing of data, and inter-organizational use of computer software and programs. This means that the instruction reportoire of the given computer should be as industry-compatible as possible, and that material on alternate choices of "third generation computers" should be taken into account during the process of evaluating future computer equipment requirements of DDC.

Once the decision has been made by DDC as to the computer equipment to be utilized from Fiscal Year 1969 onwards, the next type of decision to be made is selection of an overall "family" or grouping of computer software aids to be procured along with such equipment. The point to be made here is that a general-purpose, widely used, and industry-compatible software (programming) language should be selected for implementation of computer programs upon the selected computer equipment complex to be installed at DDC. It has been tentatively suggested in the report body that the DOD-recommended COBOL language be considered for selection in association with any computer which may be selected for installation in Fiscal Year 1969. Whatever the overall computer software language selected, it should have a comprehensive grouping of compiler, assembly, and utility "packages" (systems-level computer programming tools) which allow working, operational computer programs to be efficiently developed in a multiple-installation environment.

Once this basic selection has been made, it is then possible to talk of DDC development of computer software, programs and aids for long-term usage within a multiple-installation and multiple-application environment. Use of more than one such system or grouping of computer program languages makes this a difficult task. This is because programs written by different installations within the same operating service or environment will have to be "reprogrammed" in order that they be utilized in installations using different computer languages, which is a most expensive and time-consuming task. The other reason for this approach of one overall computer language

system for use within a DDC-user network is because of the scarcity and difficulty in obtaining highly skilled professional programming and computer systems personnel, whose talents optimally will not be utilized in clerical and redundant processes exemplified by unnecessary reprogramming. This is particularly true in the case of the multiple-organization user groups who receive a large portion of DDC products and services -- their ability to obtain the requisite computer programming and computer systems personnel is widely variant due to such factors as size, financial resources, and overall scientific and technical commitments within the organization. This means that their needs for efficient use of scarce programming and computer systems personnel must be considered as a primary factor in the selection of an overall computer language system for DDC.

Once the selection of a computer language system is completed, there is a primary research and development activity which should be initiated immediately. This is in keeping with the generic flow of materials through the Defense Documentation Center whether of management data information or of scientific and technical documentation. This is a development of general-purpose computer programs for the processing of incoming materials. This will conclude the ability to display, using computer time-sharing techniques, upon a computer-associated console of incoming machine readable materials for DDC processing. Required elements of this capability are preparation of machine readable equivalents of incoming documentary materials; provision of devices for reading such machine readable materials

into the storage units of the computer, and development of computer programs for detecting routine clerical errors contained within the various types of documentary input materials received at DDC. These latter programs are envisioned as containing directories, which will allow the testing and validation of discrete elements of information found within various types of input materials. This general-purpose capability is not present in existing DDC programs, which only may operate upon fixed, single form, previously-defined formats for each of the two major classes of inputs currently machine-processed at DDC. The provision of directories accompanying the machine-readable inputs will allow general-purpose programs to operate upon more than one set of input data, rather than use of a separate computer program for each class of individual type of input being checked. This is an existing computer capability, most commonly found in real time applications -- it should be developed at DDC.

Once this type of program has been developed, there is a call for additional software to accomplish another general-purpose function. This is compilation of comprehensive statistical information on the nature, accuracy, and utility of incoming documentary materials within the various classes utilized by DDC in storing, processing, and distributing such materials. This would apply throughout the processing cycle of DDC. What is entailed here is development of a series of computer program subroutines capable of being inserted into various DDC computer programs used in the processing of DDC materials. Assuming that time-shared computer operation will be developed to assist DDC professional processing of documentary inputs,

these subroutines could be used to calculate ongoing values for the number of documents accepted from a given source organization; number of documents rejected; number of documents from a source organization completing the entire processing cycle at DDC; number of documents being requested by users from within this accepted group; number of such documents being included in such DDC publications at the TAB, and other alike operational and management data.

A requirement for successful use of documentation across interrelated multiple organizations is the making requests to different elements of such a group and obtaining comparable retrieval responses from similar separate collections. This being the case, two types of software development should be undertaken by the Defense Documentation Center. The first is a set of computer programs designed to aid vocabulary and thesaurus personnel of the Center in managing multiple-organization thesauri, dictionaries, word lists, lists of synonms and other system tools involved in the DDC processes of classification, descriptive cataloging, indexing, and user request formulation, retrieval, and user feedback measurement. Another set of computer programs (actually within the previous grouping or complimentary to it) would afford a capability for analysing utilization of various vocabulary tools across installations and across different files of data. It is envisioned that DDC would utilize both sets of vocabulary tool computer programs across different collections. This would enable a request from a user to be applied against more than one semiautomated, computer-stored collection in one operation of a computer-

centered retrieval cycle. This is in contrast to the current (in general) practice of executing a complete set of computer programs against each separate file of data in response to a given user request which requires information from different files at the facility. Provision of such programs would lead to marked reduction in computer time requirements and in the necessity for writing special purpose programs tailored to needs of individual users.

This capability should next be extended to the use of more than one vocabulary or thesaurus (as applied to all operational computer-based DDC information stores). This is in recognition of the activities of the current Project LEX, a Department of Defense operation aimed at providing a common thesaurus for all DOD scientific and technical documentation processing. This type of software to be developed by DDC would enable the usage of more than one such thesaurus or vocabulary in processing of a given user request. If more than one vocabulary and/or thesaurus is to be used at DDC, covering, for example, different project-discipline-oriented areas, it would be most desirable to have a set of programs allowing concurrent usage of these several thesauri or vocabularies in automated processing of a given request which crosses project or discipline area boundaries. A necessity here would be computer software tools allowing the professional vocabulary and thesaurus maintenance staff at DDC to analyze thesauri and vocabularies using professional techniques, in order that they might be comparable in structure and form, and that the computer programs fulfilling user requests be able to call upon, and rationally use,

these differing tools in the activity of satisfying user retrieval requests through a single search of multiple DDC files.

In addition to the above classes of software, it should be a projected task at DDC to develop another group of software programs. These would be programs allowing the formatting of virtually "universal" output reports for users in response to their information service requests of DDC. These should allow inclusion in machine-processed reports of documentary materials which have been appropriately extracted and edited by computer programs; inclusion of management data and statistics from management-type files and of other materials and values considered responsive to a given user request processed partially by computer equipment.

An additional capability might be used in this area. This would be computer linkage with automated microform storage equipment for "on-line", or automatic, retrieval of graphic or non-textual materials from a microform store, followed by their inclusion within general purpose output reports. This would require software to manage the microform-computer equipment operation; to select appropriate microform images; to change them in form and nature to make their included data transferrable for purposes of proper insertion of this non-textual material into an output form suitable for printing and reproduction. An advanced application of this might be conversion of such non-textual material from microform media into a form suitable for transmission over remote telecommunication channels in association with textual materials. This application is an existing technique within the state of

the art, but it has not yet applied (in any known case) to a documentation center of the nature of DDC.

There have been allusions in this Appendix to the use of time-shared computer capabilities during the internal operating cycle of the processing of incoming materials by DDC from the point of acquisition to release for external distribution. This is another major area for software development: viz., the developing of a time-shared computer system capability, utilizing the general computer software language system selected for the integrated internal DDC system. Reference here is to time-sharing as it would be used internally within the Defense Documentation Center during the discrete professional processes of acquisition, selection, classification, descriptive cataloging, indexing, editing, and publication composition. No known existing software system is available for use through such a generic cycle of preparation of secondary publications.

The Defense Documentation Center should carry on development with the philosophy that, when a general purpose system or elements thereof are developed outside DDC and its user network, DDC will consider incorporation of any such proven software aids into their own operating system, provided that they are in the chosen computer system language and compatible with system operations. DDC should move, in cooperation with other Federal agencies and related and interacting non-Federal organizations, to develop standards for the development of computer software in all the areas mentioned above. These standards also should be developed for another generic area,

that of remote transmission of documentary or report information outside the physical confines of the Defense Documentation Center. Future development of economically-feasible communication capability would make standardization of telecommunication software and procedures a primary area for DDC activity.

In company with development of software for checking the validity of incoming materials, DDC should develop computer software for optimized processing of textual and non-textual documentary material through such devices as optical scanners, page readers, and print readers. These devices, currently in a state of advanced development, require computer-type software for their efficient utilization. DDC should develop software for these equipments as well as for computer, microform, and telecommunication devices, so that the internal integrated system may eventually become capable of handling multiple files of stored materials in a network encompassing a varied group of external users and more than one government agency.

Current (and projected) DDC software development does encompass more than one government agency. An example is the activity of the Joint Interagency Operations Group, which includes AEC, DOD, NASA, and CFSTI. This group is concerned with problems of referral of retrieval requests to the appropriate agency, optimization of formats for transfer of machine readable materials between different participating agencies, and with problems of producing interagency management data and documentation files, products, and processing tools. This particular example illustrates a more general proposition,

that DDC should both participate in, and encourage, development of software tools applicable to various areas of inter-organizational activity. An example of this would be previously mentioned computer software advocated for use in the case where more than one vocabulary or thesaurus would exist in DDC for application against several files -- this applies, by extension, in the case where more than one vocabulary or thesaurus is used among different agencies for processing of user requests against similar files within the separate agencies.

These software developments for DDC use should be carried out under DDC direction. Some of them could be carried out by DDC itself; others by a combined group of DDC and its user organizations; still others by a combination of DDC and its contractors, and others by a grouping of inter-agency personnel (with or without contractor assistance). No attempt is made here to designate which effort should be produced by these various groupings. However, it is obvious in some cases that various group combinations are necessary (e.g., in the case of development of telecommunication-related software, where DDC personnel and external users or government agency organizations sharing telecommunication facilities might well be participants).

3. AREAS FOR DDC COMPUTER EQUIPMENT RESEARCH AND DEVELOPMENT

The Defense Documentation Center currently has over 300,000 scientific and technical document references stored in machine readable form, and an additional 25,000 management data records. This is believed to be the

largest existing store of machine readable scientific and technical documentation materials. This being the case, the Defense Documentation Center has a high stake in developing computer software, and in three areas of equipment development which should be of particular concern to DDC during the next five years and beyond.

The first of these three areas is that of development of more sophisticated and more effective automated microform storage and conversion media. The currently existing files of microform-stored data at DDC are in microfiche form, in accordance with DOD-COSATI standards. Parts of the store are in 16 mm. and 35 mm. microfilm form, but are being converted to the above mentioned microfiche format. These microform media are not now directly amenable to computer-directed manipulation. An example of an existing machine which affords a limited capability here is the Stromberg-Carlson 4000 Series of microfilm equipment. This type of equipment effectively allows a computer-produced magnetic tape of numeric input to be converted into graphic form, utilizing the tracing of an image through generation of points and lines upon a cathode-ray-tube; this cathode-ray-tube representation then being converted into microfilm form. This equipment also has the capability of taking a picture of this cathode-ray-tube image, thus allowing simultaneous reproduction in hardcopy form of the image which has been formed on the cathode-ray-tube and placed on microfilm. There is no known current capability for direct, automated preparation of microfiche using this type of equipment, nor is such preparation possible on any other form of equipment in use today in an operating integrated total

system. A related form of contemporary equipment is represented by the IBM Cypress, and the earlier Walnut device. These devices effectively allow a digital computer to manage the formation, storage and retrieval of microform images of textual and non-textual materials. Another existing class of microform-handling equipment is that of microfiche and microfilm readers and printers. The devices are currently in an advanced state of development, and are not, insofar as is known, used in coordination with operational computer-associated consoles used for display for textual materials.

Considering the size of the DDC files as they exist and their increased size as projected into the future, the DDC should undertake a strong role in development of computer-directed microform creation, manipulation and retrieval devices. In addition, DDC should strongly promote development of computer-associated consoles which would allow both display of graphic (primarily non-textual) images obtained from such computer-directed microform devices, and of textual materials obtained from computer storage. These areas of equipment development should be paralleled by DDC fostering of appropriate computer and microform equipment software capable of allowing efficient utilization of the aforementioned types of equipment in an overall time-shared-multiple-mode, computer-centered system capable of continuing integrated assistance in the methods utilized in DDC processing of incoming materials, of external user requests, and of interagency products and services.

A second major area for DDC development is that of research and telecommunication-supporting software. In association with this, DDC should promote development of more rapid and more economical equipment allowing medium speed transmission of graphic images across telecommunication channels. The maximum existing (unclassified) applications of such transmission media allow approximately 15 typewritten pages of text to be transmitted across medium speed transmission channels per minute of operation. The cost of equivalent, current service will not be exemplified here: sufficient to say that it is quite expensive. DDC should promote more economical availability of transmission channels per se, and the development of general-purpose, high capacity, and economically reasonable graphic image facsimile transmission and reproduction terminals. These terminals should be capable of being linked within a computer-microform-composition equipment-telecommunication network of equipment (and related software) serving needs of the Defense Documentation Center, its user groups, and other government agencies in non-Federal organizations with interrelated and common objectives.

The third equipment class wherein DDC should assume a leading development role is that of high-capacity equipment capable of composing graphic (i.e., non-textual) and textual materials into complete documentary output products suitable for publication. No known operational equipment has this capability today. The CBS Laboratories-Mergenthaler Corporation Linotron device is an advanced, computer-operated photographic composition machine in use at the Government Printing Office. It represents the most advanced known

composition device, but it does not have this type of capability of allowing semi-automated incorporation of graphic images into composed text to form a total document image in complete form for printing. There is still a requirement, even in the case of the Linotron, for preparation of "mockups" and of "galleys", which contain textual material which must be augmented with graphic images, in a manual process, to form the final product for printing. This appears to be an area where joint interagency and inter-Federal branch cooperation might be beneficial both to the Government Printing Office and to various agencies in the Federal Executive Branch.

A final area for development which appears to be strongly pertinent to DDC's future is that of high-quality reproduction devices which allow material to be printed at a high rate of speed, the fastest known being the Radiation, Incorporated printer which produces 30,000 pages of text per hour. However, this printer utilizes "flimsy", the teletype or facsimile style of paper, which is not suitable for publication of the TAB or equivalent products. There is need for faster reproduction of outputs in response to users services of DDC which are prepared through the use of computer equipment -- this is an area in which availability of advanced equipment would be of high value to DDC in its future operations. This implies a requirement for faster printing of output for foreseeably larger user groups, which is currently beyond the effective capabilities of standard computer printers.

These equipment development areas are felt to be particularly pertinent to DDC's present and foreseeable requirements, and represent development areas

which have lagged behind other areas (e.g., of computer memories, internal computer operating speeds, etc.). DDC might benefit by promoting computer development of more rapid computer input/output devices, principally by magnetic tape, which still represents the major form of storage for documentation at most existing installations.

Many of the cited development applications normally are undertaken by manufacturers of the particular equipment. This is true of the hardware or equipment, while software and programs more typically are developed by a combination of manufacturer and installation personnel activities. DDC should foster manufacturer developments of the equipment per se; of as much software as the equipment manufacturer can or is willing to provide, and of DDC and contractor preparation of additional software. There are cases (as cited above) where interagency cooperation and development of software would be very advisable.